FCSM 2018: Mapping Geographic and Temporal Variations in select Natality and Mortality Outcomes with R-INLA in Small Areas

Diba Khan
Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Research and Methodology
Hyattsville, MD.
Disclaimer

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official positions of the National Center for Health Statistics, Centers for Disease Control and Prevention.
Co-authors and acknowledgement

1. Teen birth rates and visualization: Brady Hamilton, Lauren Rossen, Rong Wei, Yulei He, and Yinong Chong
2. Suicides mortality: Lauren Rossen, Holly Hedegaard, and Margaret Warner
3. Acknowledgement: Makram Talih and Jessica Keralis
Objective

To explore geographic and temporal variation in natality (teen births) and mortality (suicides) rates at the county level in U.S. using the National Vital Statistics data for the years 2003-2015 and 2005-2015 respectively.
Outline

1. Bayesian philosophy
2. Integrated Nested Laplace Approximation
3. Example: Teen birth rates
4. Space time models
5. Model check and accuracy
6. Results
7. Data visualization
8. Example: Suicides mortality
9. Space time models
10. Model check and accuracy
11. Results
12. Conclusions
Hierarchical Bayesian Model

1. Data Likelihood: \( L(\mathbf{y}|\text{parameters}) \), where \( \mathbf{y} = (y_1, y_2, \ldots, y_m) \), \( i=1,\ldots,m \) areas

2. Prior distribution for the model parameters: \( \text{Prob}(\text{parameters}) \)

3. Prior distributions: uninformative or vague priors (are not assumption free)

4. Jeffreys priors are invariant to linear transformations but are improper

5. **Posterior distribution:**
   \[
   \text{Prob}(\text{parameters}|\mathbf{y}) \propto L(\mathbf{y}|\text{parameters}) \cdot \text{Prob}(\text{parameters})
   \]

6. Hierarchical Bayes: an extra level of hierarchy in setting the prior distribution of model parameters

7. Random effects: extra variability due to unmeasured confounders modeled by assigning an individual units effect
Flat priors

**Figure:** Priors.
Spatially correlated random effects

1. Spatially correlated effects: Intrinsic Conditional Autoregressive (ICAR) prior distribution
2. Spatial locations for areal data
3. Esri shapefile - polygon file gives geographical coordinates of the boundaries of each area
4. Weights are used to express spatial dependence between areas
5. Most commonly used specifications of weights: binary specification
6. The conditional expectation of ICAR prior random effect for an area is the average of the random effects in neighboring areas
7. The conditional variance of ICAR prior random effect for an area is inversely proportional to the number of neighbors
Posterior distribution: INLA, Winbugs, STAN, JAGS

1 Traditionally posterior distribution is estimated via Markov Chain Monte Carlo (MCMC): exact inference, extremely flexible - applicable to any type of data and model, computational and time intensive

2 MCMC: Fundamental issues - model complexity and database dimension

3 Gibbs sampling and Metropolis-Hastings available in WinBugs, OpenBUGS, JAGS

4 Posterior distribution can also be approximated via Laplace approximation in R-Integrated Nested Laplace Approximation (INLA) package: computationally efficient alternative to MCMC, reliable estimates in less time, particularly relevant to large datasets

5 Less established as compared to MCMC
R-INLA

1. Approximates the posterior marginals of a variety of Bayesian hierarchical models
2. Linear Mixed Models or Generalized Linear Mixed Models, Spatial, and Spatio-Temporal models, and point process or Geostatistical models
3. Approximation via Laplace integral approximation to the fixed effects
4. Numerical integration approximation to the random effects
5. **Exceedance probability:** allows faster computation of the posterior probability that a parameter does/does not exceed a certain threshold
Small area direct estimates: example of county level teen births for the age group: 15-19 for the years 2003-2015
Data suppression: Births less than 20 suppressed, 15-19, 2015

Figure: Teen birth rates based on less than 20 births for a county are suppressed for the year 2015.
Total number of counties by counts of teen births and percentage in the age group 15 – 19 for years 2003-2015

<table>
<thead>
<tr>
<th>Year</th>
<th>Equal to 0</th>
<th></th>
<th>Less than 10</th>
<th></th>
<th>Less than 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
</tr>
<tr>
<td>2003</td>
<td>56</td>
<td>1.78</td>
<td>512</td>
<td>16.31</td>
<td>929</td>
</tr>
<tr>
<td>2008</td>
<td>54</td>
<td>1.72</td>
<td>516</td>
<td>16.44</td>
<td>880</td>
</tr>
<tr>
<td>2015</td>
<td>102</td>
<td>3.25</td>
<td>797</td>
<td>25.4</td>
<td>1403</td>
</tr>
</tbody>
</table>

Table: Counts and percentage for total number of counties by counts of teen births in the age group 15 – 19 reported to be equal to 0, less than 10, and less than 20 for 2003, 2008, and 2015 respectively.
**Why Bayesian approach?**

1. Main advantage: takes into account the uncertainty of the estimates/predictions
2. The inferential process accounts for spatial trend via spatially structured random effects: providing insight knowledge
3. Hierarchial model accounts for similarities based on neighbourhood structure
4. The space-time effects explain the differences/changes in time trend for different counties
Bayesian model

- Define a spatial-temporal model using a hierarchical Bayesian framework
- Account for spatial and temporal trends
- Areas close to each other - more likely to share geographical characteristics related to the health outcome
- Identification of temporal pattern: stronger for subsequent years than for years apart
Covariates: Principal Component Analysis (PCA)

- Calculated the correlation between continuous covariates and the teen birth rate each year from 2003 - 2015
- Selected the covariates for which the absolute correlation was greater than 0.4 for a majority of the years
- There were 18 covariates selected
- Conducted PCA on these covariates
- Factors with an eigen value greater than 1 were retained, in this case 3 factors
- Varimax rotation was used
PCA - Construct 1 - High poverty and low income
PCA - Construct 2 - Educational level
PCA - Construct 3 - Race/Ethnicity: Percent White
Problem formulation

\( y_{it} \) : counts of births by county \( i \) and year \( t \)
\( n_{it} \) : counts of population by county \( i \) and year \( t \)
\( y_{it} \sim Binomial(n_{it}, p_{it}) \)
\( p_{it} \) : probability of teen births in county \( i \) at time \( t \)

where \( i = 1, \ldots, m \) areas and \( t = 1, \ldots, T \) times.

General space-time model structure (Lawson, A. (2013)):
\[
\logit(p_{it}) = \alpha_0 + A_i + B_t + C_{it},
\]

where:
\( A_i \) : spatial group
\( B_t \) : temporal group
\( C_{it} \) : space-time interaction group
Proposed models

**Model 1** With uncorrelated heterogeneity (non-spatial random effect)

\[
\text{logit}(p_{it}) = \alpha_0 + v_i
\]

Prior for \(\alpha_0 \sim \text{dflat}\)

Prior for \(v_i\): \(v_i \sim N(0, 1/\tau_v)\) termed as uncorrelated heterogeneity (variability)

\(\tau_v\) is the precision

\[\text{Log}(\tau_v) \sim \text{LogGamma}(1, 0.001)\]
Proposed models

Model 2 Besag: convolution model

\[ \logit(p_{it}) = \alpha_0 + u_i + v_i \]

Intrinsic CAR prior for: \( u_i \mid u_{-i} \)

\[ u_i \mid u_{-i} \sim N(\bar{u}_i, r/n_{\delta_i}) \] termed as correlated heterogeneity (variability), where

\[ u_{-i} = (u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m) \]

\( n_{\delta_i} \) : number of neighbors = \( \sum_{j=1}^{m} \omega_{ij} \)

\( \delta_i \) : neighborhood of \( i^{th} \) region

\( r \) : is the variance, \( r = 1/\tau_r : \tau_r \) is the precision

\[ \log(\tau_r) \sim \log\text{Gamma}(1, 0.001) \]

\[ \bar{u}_i = \frac{\sum_{j=1}^{m} \omega_{ij} u_j}{\sum_{j=1}^{m} \omega_{ij}} \]

\( \alpha_0 \sim dflat \)
Proposed models

**Model 3** Random time effect (uncorrelated)

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma_1 t \\
\gamma_2 t \sim N(0, 1/\tau_{\gamma_2}) \text{ (uncorrelated)} \\
\log(\tau_{\gamma_2}) \sim \text{LogGamma}(1, 0.001)
\]
Proposed models

**Model 4** Random time effect (uncorrelated) plus iid space-time interaction

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma_2 t + \psi_{it}
\]

\[
\gamma_2 t \sim N(0, 1/\tau_{\gamma_2}) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_{\gamma_2}) \sim \text{LogGamma}(1, 0.001)
\]

\[
\psi_{i,t} \sim N(0, 1/\tau_{\psi}) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_{\psi}) \sim \text{LogGamma}(1, 0.001)
\]
Proposed models

**Model 5** Space-time interaction (correlated)

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma_2 t + \psi_{it}
\]

\[
\gamma_2 t \sim N(0, 1/\tau_{\gamma_2}) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_{\gamma_2}) \sim \text{LogGamma}(1, 0.001)
\]

\[
\psi_{i,t} \sim N(\psi_{i,t-1}, \tau_{\psi}) \text{ (randomwalk (Type 2 interaction))}
\]

\[
\text{Log}(\tau_{\psi}) \sim \text{LogGamma}(1, 0.001)
\]
Proposed models

**Model 6** Random time effect (uncorrelated) plus space-time interaction (correlated) plus covariates

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma^{2t} + \psi_{it} + \mathbf{X}'\beta
\]

\[
\gamma^{2t} \sim N(0, 1/\tau_{\gamma^2}) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_{\gamma^2}) \sim \text{LogGamma}(1, 0.001)
\]

\[
\psi_{i,t} \sim N(\psi_{i,t-1}, \tau_{\psi}) \text{ (randomwalk (Type 2 interaction))}
\]

\[
\text{Log}(\tau_{\psi}) \sim \text{LogGamma}(1, 0.001)
\]

\[
\mathbf{X} \text{ : covariates matrix } \mathbf{X}_i
\]

\[
\beta \text{ : vector of regression parameters}
\]

\[
\beta \sim N(0, 100)
\]
Goodness of the fit: Deviance Information Criterion (DIC) and Watanabe-Akaike information criteria (WAIC)

Table: DIC, the effective number of parameters estimated and WAIC for the models via INLA.

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>n.eff</th>
<th>WAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>472039</td>
<td>3083.18</td>
<td>489985</td>
</tr>
<tr>
<td>Model 2</td>
<td>471998.3</td>
<td>3026.625</td>
<td>489911.6</td>
</tr>
<tr>
<td>Model 3</td>
<td>285111.8</td>
<td>3037.31</td>
<td>287390.2</td>
</tr>
<tr>
<td>Model 4</td>
<td>271376</td>
<td>12347.61</td>
<td>271407.5</td>
</tr>
<tr>
<td>Model 5</td>
<td>267375.5</td>
<td>8406.752</td>
<td>268159.3</td>
</tr>
<tr>
<td><strong>Model 6</strong></td>
<td><strong>267251.3</strong></td>
<td><strong>8684.229</strong></td>
<td><strong>267528.2</strong></td>
</tr>
</tbody>
</table>
Model check: State direct TBRs vs INLA model based aggregated state level TBRs, 2003-2015

Figure: State direct TBRs vs INLA based state TBRs 2003-2015.
Model accuracy: MCMC vs INLA TBRs. Computation time: INLA (24 hours) vs MCMC (9 weeks)

Figure: MCMC based vs INLA based aggregated state level TBRs, 2003-2012.
Predicted TBRs 2015
Animation of the predicted TBRs: 2003 through 2015
Animation of Type II interaction effects: 2003-2015
Exceedance probabilities

1. The probability that the predicted TBRs exceed a certain threshold

2. Can be used to quantify the uncertainty associated with the county level TBRs. Example: Probability that the predictions exceed the mean county level TBRs
Exceedance probabilities: threshold

Probability of exceeding a threshold
Threshold set as the mean crude teen birth rate for the year 2015
Animation of Exceedance probabilities: 2003 through 2015
Trends in TBRs: uncorrelated time effect $\gamma_{2t}$ and the 95 percent Bayesian credible intervals

Figure: Uncorrelated time effect $\gamma_{2t}$. 
Correlated heterogeneity

Figure: Correlated heterogeneity $u_j$. 
Data visualization: Trends in teen birth rates

1. Data visualization done in the software Tableau

2. State level: Interactive maps and trend lines for teen birth rates (from birth certificates) for females aged 15-19, 15-17, and 17-19 for each of the 50 states

3. Can zoom in on individual states
Data visualization: State level teen birth rates: 1990-2015
Data visualization: County level teen birth rates, 2003-2015

1. Contains four dashboards
2. County level: Interactive maps and trend lines for estimated teen birth rates for females aged 15-19 for 3137 counties
3. Trend lines, geographic variation and 95 percent Bayesian credible bands for the years 2003-2015 (soon to be updated with 2016)
4. Can zoom in on individual counties for more granular look
Teen Birth Rates for Age Group 15–19 in the United States by County, 2003-2015

These figures display estimated teen birth rates for the age group 15–19 (expressed per 1,000 females aged 15–19) by county and year in a series of interactive maps, median and trend plots, and 95% Bayesian credible bands for each county and state in the United States for the years 2003–2015. The estimated teen birth rates for the age group 15–19 were calculated using Bayesian space-time hierarchical models and explanatory variables for the observed teen birth data from the year 2003-2015 for each county and year. More information on methods can be found in the Notes below the visualization.

The first three dashboards show heat maps of estimated teen birth rates for the age group 15–19 by county and year. The first dashboard displays the continental U.S.; the second displays the Northeast Census region (in order to provide more granular look and be able to view the counties in the smaller states in more detail); and the third shows Alaska, Hawaii, and the District of Columbia, where Alaska and Hawaii are not included in the map of the continental U.S. in the first dashboard and District of Columbia is shown separately to provide a more detailed view. The color scale indicates the magnitude of the estimated county teen birth rate from lowest (light color) to highest (dark color). The county grid on the right-hand side shows the change in estimated teen birth rate by year using the same color scale as the map.

- Use the arrows or the slider to select a year. Click on any state to zoom into it on the map.
- Clicking on a state will update the list to show counties for that state. Selecting a county on the map will highlight that county in the grid.
- Click outside the state to remove the county highlight, and click again to zoom back out to the map of the continental U.S. Users may click on the gray “home” icon in the upper right-hand corner to return to the map of the continental U.S.
Data visualization


Select Year
2015

Counties in Maine
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York

Legend for estimated birth rate per 1,000 females aged 15–19
0-20
20-25
25-40
40-50
50-65
65-90
90+
Data visualization


Select Year
2015

Counties in Delaware

Legend for estimated birth rate per 1,000 females aged 15–19

- 0-20
- 20-25
- 25-40
Data visualization

Estimated Teen Birth Rates for Females Aged 15–19 by County: Alaska, Hawaii, and DC, 2015

Boroughs and Census Areas in Alaska

- Aleutians East
- Aleutians West
- Anchorage
- Bethel
- Bristol Bay
- Denali
- Dillingham
- Fairbanks North Star
- Haines
- Juneau
- Kenai Peninsula
- Kodiak Island
- Lake and Peninsula
- Matanuska Susitna
- Nome
- North Slope
- Northwest Arctic
- Prince Wales Ketchikan
- Sitka
- Skagway Hoonah Angoon
- Southeast Fairbanks
- Valdez Cordova
- Wade Hampton
- Yakutat
- Yukon Koyukuk

Counties in Hawaii

- Hawaii
- Honolulu
- Kauai
- Maui

Legend for estimated birth rate per 1,000 females aged 15–19

- 0-20
- 20-25
- 25-40
- 40-50
- 50-65
- 65-90
- 90+
Data visualization

Select a State

County Estimated Teen Birth Rate Medians and Credible Intervals

Estimated Teen Birth Rates for Females Aged 15–19 for Counties in Arkansas

Select a county from the table below to display it in the Credible Intervals line graph.

Select a year to display estimated teen birth rates and 95% Bayesian credible bands for females aged 15–19 in the above table.

2003
Select a State

Below is a pair of two line graphs showing the change in estimated teen birth rates. The Median Teen Birth Rate graph shows the change in estimated teen birth rates for the counties in the selected state from 2003 to 2015. Each county's estimated rate is displayed as a gray line, and the median estimated teen birth rate for all counties in the selected state is displayed as a blue line. The Credible Intervals graph shows the estimated teen birth rate for the selected county as a black line, and the upper and lower Bayesian credible intervals as orange lines. Year is on the X-axis, and estimated teen birth rate per one thousand is on the Y-axis.

County Estimated Teen Birth Rate Medians and Credible Intervals

Select a county from the table below to display it in the Credible Intervals line graph.

Estimated Teen Birth Rates for Females Aged 15–19 for Counties in Mississippi

Select a year to display estimated teen birth rates and 95% Bayesian credible bands for females aged 15–19 in the above table.

2003

<table>
<thead>
<tr>
<th>County</th>
<th>Estimated Teen Birth Rate</th>
<th>Lower Confidence Limit</th>
<th>Upper Confidence Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarke</td>
<td>58.8</td>
<td>50.1</td>
<td>68.4</td>
</tr>
<tr>
<td>Clay</td>
<td>60.4</td>
<td>52.2</td>
<td>69.2</td>
</tr>
<tr>
<td>Coahoma</td>
<td>91.2</td>
<td>81.7</td>
<td>101.3</td>
</tr>
<tr>
<td>Copiah</td>
<td>60.9</td>
<td>53.7</td>
<td>68.7</td>
</tr>
<tr>
<td>Covington</td>
<td>73.8</td>
<td>64.0</td>
<td>83.9</td>
</tr>
<tr>
<td>De Soto</td>
<td>46.6</td>
<td>42.2</td>
<td>51.1</td>
</tr>
<tr>
<td>Forrest</td>
<td>48.9</td>
<td>44.1</td>
<td>53.9</td>
</tr>
<tr>
<td>Franklin</td>
<td>50.5</td>
<td>41.1</td>
<td>61.2</td>
</tr>
<tr>
<td>George</td>
<td>84.1</td>
<td>73.3</td>
<td>95.8</td>
</tr>
<tr>
<td>Greene</td>
<td>61.0</td>
<td>50.9</td>
<td>72.2</td>
</tr>
<tr>
<td>Grenada</td>
<td>65.1</td>
<td>56.5</td>
<td>74.5</td>
</tr>
<tr>
<td>Hancock</td>
<td>53.3</td>
<td>47.0</td>
<td>60.2</td>
</tr>
<tr>
<td>Harrison</td>
<td>64.1</td>
<td>59.5</td>
<td>68.8</td>
</tr>
<tr>
<td>Hinds</td>
<td>60.1</td>
<td>56.3</td>
<td>64.0</td>
</tr>
<tr>
<td>Holmes</td>
<td>79.3</td>
<td>69.9</td>
<td>89.4</td>
</tr>
<tr>
<td>Humphreys</td>
<td>103.6</td>
<td>89.6</td>
<td>118.6</td>
</tr>
<tr>
<td>Issaquena</td>
<td>74.2</td>
<td>53.9</td>
<td>97.8</td>
</tr>
<tr>
<td>Itawamba</td>
<td>48.9</td>
<td>41.9</td>
<td>56.4</td>
</tr>
</tbody>
</table>
Data visualization
Data visualization

Data visualization

Figure: Estimated median teen birth rates (per 1,000) for counties in Alabama.
Small area direct estimates: example of county level suicide rates (SRs)

1. Age adjusted suicide rates increased by 27 percent from 1999-2015 (10.5-13.3 per 100,000 population)
2. County level estimates for less than 20 events are suppressed
3. Unreliable: for example rural areas fewer suicides and small population sizes
4. Past studies aggregate over several years (for example WISQARS (2008-2014)) may mask temporal trends
5. Aggregation of counties to produce larger geographic areas: may mask urban-rural differences and sub-state variations
Total number of counties by counts of suicides for years 2005-2015

Table: Counts and percentage for total number of counties by counts of suicides reported to be equal to 0, less than 10, and less than 20 for each 2005, 2009, and 2015 respectively.

<table>
<thead>
<tr>
<th>Year</th>
<th>Equal to 0</th>
<th>Less than 10</th>
<th>Less than 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percent</td>
<td>Count</td>
</tr>
<tr>
<td>2005</td>
<td>475</td>
<td>15.12</td>
<td>2405</td>
</tr>
<tr>
<td>2009</td>
<td>427</td>
<td>13.6</td>
<td>2349</td>
</tr>
<tr>
<td>2015</td>
<td>360</td>
<td>11.5</td>
<td>2186</td>
</tr>
</tbody>
</table>
WISQARS: aggregation in time (2008-2014)

**Figure:** Geographic variation for aggregated suicide rates over 7 years.
Data suppression: Crude suicide number of deaths for the year 2015

Figure: Suicide rates based on less than 20 suicides for a county are suppressed for the year 2015.
Why Bayesian approach?

1. Main advantage: takes into account the uncertainty of the estimates/predictions
2. The inferential process accounts for spatial trend via spatially structured random effects: providing insight knowledge
3. Hierarchical structure accounts for similarities based on neighbourhood structure
4. The space-time effects explain the differences/changes in time trend for different counties
Problem formulation

\[ y_{it} : \text{counts of suicides by county } i \text{ and year } t \]
\[ n_{it} : \text{counts of population by county } i \text{ and year } t \]
\[ y_{it} \sim Binomial(n_{it}, p_{it}) \]
\[ p_{it} : \text{probability of suicides in county } i \text{ at time } t \]

where \( i = 1, \ldots, m \) areas and \( t = 1, \ldots, T \) times.

General space-time model structure (Lawson, A. (2013)):

\[ \text{logit}(p_{it}) = \alpha_0 + A_i + B_t + C_{it}, \]

where:

\( A_i \) : spatial group
\( B_t \) : temporal group
\( C_{it} \) : space-time interaction group
Proposed models

**Model 1** With uncorrelated heterogeneity (non-spatial random effect)

\[ \text{logit}(p_{it}) = \alpha_0 + v_i \]

Prior for \( \alpha_0 \sim \text{dflat} \)

Prior for \( v_i \): \( v_i \sim N(0, 1/\tau_v) \) termed as uncorrelated heterogeneity (variability)

\( \tau_v \) is the precision

\[ \log(\tau_v) \sim \text{LogGamma}(1, 0.001) \]
Proposed models

**Model 2** Besag: convolution model

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i
\]

Intrinsic CAR prior for: \(u_i|u_{-i}\)

\[
u_i|u_{-i} \sim N(\bar{u}_i; r/n_{\delta_i}) \text{ termed as correlated heterogeneity (variability), where}
\]

\[
u_{-i} = (u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m)
\]

\[n_{\delta_i} : \text{number of neighbors} = \sum_{j=1}^m \omega_{ij}\]

\[\delta_i : \text{neighborhood of } i^{th} \text{ region}\]

\[r : \text{is the variance, } r = 1/\tau_r: \tau_r \text{ is the precision}\]

\[\text{Log}(\tau_r) \sim \text{LogGamma}(1, 0.001)\]

\[\bar{u}_i = \frac{\sum_{j=1}^m \omega_{ij} u_j}{\sum_{j=1}^m \omega_{ij}}\]

\[\alpha_0 \sim dflat\]
Proposed models

**Model 3** Random time effect (correlated)

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma_{1t}
\]

\[
\gamma_{1t} \sim N(\gamma_{1t-1}, 1/\tau_{\gamma_1}) \text{ (randomwalk)}
\]

\[
\log(\tau_{\gamma_1}) \sim \text{LogGamma}(1, 0.001)
\]
Proposed models

**Model 4** Random time effect (uncorrelated)

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma_2 t
\]

\[
\gamma_2 t \sim N(0, 1/\tau_{\gamma_2}) \quad \text{(uncorrelated)}
\]

\[
\log(\tau_{\gamma_2}) \sim \text{LogGamma}(1, 0.001)
\]
Proposed models

**Model 5** Space-time interaction (uncorrelated)

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t + \psi_{it}
\]

\[
\psi_{i,t} \sim N(0, 1/\tau_\psi) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_\psi) \sim \text{LogGamma}(1, 0.001)
\]
Covariates

1. Covariates enhance small area predictions
2. 53 variables identified based on past literature studies
3. Demographic variables, health-related characteristics, socioeconomic factors, treatment gap for drug and alcohol use, county level model based estimates of drug poisoning
Proposed models

**Model 6** Model 5 plus covariates

\[
\text{logit}(p_{it}) = \alpha_0 + u_i + v_i + \gamma 1_t + \psi_{it} + X'\beta
\]

\[
\psi_{i,t} \sim N(0, \tau_\psi) \text{ (uncorrelated)}
\]

\[
\text{Log}(\tau_\psi) \sim \text{LogGamma}(1, 0.001)
\]

\(X\) : time varying covariates matrix \(X_{it}\)

\(\beta\) : vector of regression parameters

\(\beta \sim N(0, 100)\)
Goodness of the fit: Deviance Information Criterion (DIC) and Watanabe-Akaike information criteria (WAIC)

Table: DIC, the effective number of parameters estimated and WAIC for the models via INLA.

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>n.eff</th>
<th>WAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>150371.4</td>
<td>2316</td>
<td>150763.3</td>
</tr>
<tr>
<td>Model 2</td>
<td>149966.2</td>
<td>2316</td>
<td>150280.2</td>
</tr>
<tr>
<td>Model 3</td>
<td>148008.6</td>
<td>1884</td>
<td>148166.2</td>
</tr>
<tr>
<td>Model 4</td>
<td>148010.3</td>
<td>1886</td>
<td>148168.1</td>
</tr>
<tr>
<td>Model 5</td>
<td>147821.9</td>
<td>2766</td>
<td>147938</td>
</tr>
<tr>
<td>Model 5 + covs</td>
<td>147181.1</td>
<td>1896</td>
<td>147250</td>
</tr>
</tbody>
</table>
Model accuracy: State direct Suicide Rates vs INLA model based aggregated state level Suicide Rates, 2005-2015

Figure: State direct SRs vs INLA based state SRs 2005-2015.
Shrinkage of model based Suicide Rates for each state, by population size for 2015

**Figure:** Shrinkage of suicide rates for each state, by population size for 2015. Crude death rates are plotted at the start of the arrows, and model-based death rates are located at the end of the arrows.
Model accuracy: MCMC SRs vs INLA SRs, Computation time: MCMC: (8 weeks) vs INLA (24-36 hours)

Figure: MCMC based vs INLA based aggregated state level Suicide Rates, 2005-2015.
Predicted Suicide Rates (SRs) 2005

Figure: Predicted posteriors 2005.
Predicted Suicide Rates (SRs) 2015

Figure: Predicted posteriors 2015.
Animation of Suicide Rates: 2005 through 2015
Covariates significantly associated with SRs

1. Demographic characteristics: household size, racial and ethnic distribution, urbanization level, and divorce rates
2. Socioeconomic factors: median home value, median gross rent, household crowding, and median per capita income, percent persons with college education, unemployment rate, high-cost loan rate
3. Health-related characteristics: percent abusing or dependent on illicit drugs or alcohol in the previous year, treatment gap for alcohol and drug use, and prevalence of major depressive episode
4. County-level model-based estimates of age-adjusted death rates due to drug poisoning
5. Consistent with prior analyses reporting county-level (i.e., ecological) associations between socioeconomic, demographic and/or health-related factors and suicide rates
Exceedance probabilities

1. The probability that the predicted SRs exceed a certain threshold

2. Can be used to quantify the uncertainty associated with the county level SRs. Example: Probability that the predictions exceed the mean county level SRs
Uncertainty associated with the county level SRs, year 2005

Figure: Predicted county level SRs and the probability of exceeding the crude mean county level SR for the year 2005 (14.61 per 100,000).
Uncertainty associated with the county level SRs, year 2015

Figure: Predicted county level SRs and the probability of exceeding the crude mean county level SR for the year 2015 (18.74 per 100,000).
Healthy People 2020 (HP2020) sets targets

HP2020 uses the age-adjusted suicide rate for 2007 for state level SRs, which is 11.3 per 100,000 to set targets. Apply a 10 percent improvement to get the target of 10.2 per 100,000

For crude county level suicide rates: mean crude rate for 2007 is 14.91 per 100,000. Apply a 10 percent improvement to get the target of 13.419 per 100,000
State-level Data 2015: Suicide rate (age adjusted, per 100,000 population)

Pr (exceeding the target rate of 0.00013419) - 2005-2015.
Trends in suicide rates: correlated time effect, Type II random walk, $\gamma_{2t}$ and 95 percent Bayesian credible intervals.

**Figure:** Correlated time effect $\gamma_{2t}$. 
Uncorrelated heterogeneity and correlated heterogeneity

Figure: Uncorrelated heterogeneity $v_i$ and correlated heterogeneity $u_i$. 
Conclusions

1. Use of R-INLA method resulted in substantially reduced computation time: 8 weeks vs 24 hours
2. A variety of time and random effects could be tested
3. R-INLA Allows faster computation of exceedance probabilities to determine if counties have met the specified targets/thresholds
4. The functional form of the covariates in R-INLA can be specified in different forms
Conclusions

1. TBRs declined across all regions of the country from 2003 to 2015
2. TBRs remained in excess of 67 births per 1,000 adolescent females in several counties across Texas, along the Mississippi river, Montana, New Mexico, Georgia and Alaska
3. Higher TBRs across counties in the southern U.S. and lower TBRs in New England counties during the study period, 2003 to 2015
4. Large variation in TBRs in smaller counties within states and large teen birth rates in rural areas
5. Data visualization: can zoom in on counties for more granular look
6. Data visualization: identify counties in greatest need
Conclusions

1. All counties demonstrated an increase in suicide death rates from 2005-2015
2. The counties with the highest suicide rates were predominantly located across the western US
3. The counties with lowest rates were observed across southern California, western Texas, along the Mississippi river, and in areas along the East Coast
4. Several county-level covariates, namely socioeconomic, demographic, and/or health related factors were found to be significant predictors of SRs
Future research

1. Smoothed county level estimates can be used to assess urban-rural disparities
2. Future research examining spatiotemporal trends by age and gender would be informative
3. Spatial clustering at the sub county levels would provide additional insights
4. Future research can look at county-level variation by race and Hispanic origin groups
5. Future research on Neonatal intensive care units (NICU) births at the county level would be informative
Ongoing research: Hot and cold spots for Teen Birth Rates, females aged 15-19, 2016 for Hispanics

**Figure:** Values represent z-scores from the Getis Ord Gi analysis via ArcGis. Negative z-scores indicate cold spots, while positive z-scores indicate hot spots.
References


References


References

1 Rue, H. and Martino, S. INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. 2009. R package, version 0.0.


3 http://www.r-inla.org/

Contact information

1. Diba Khan ild1@cdc.gov or 301-458-4474
2. For teen births visualization: Brady E. Hamilton boh5@cdc.gov or 301-458-4653